mother-cell proteins to the sporulation septum, a physical uniqueness that distinguishes the septal membrane from other regions of the cell seems to have been discovered. After insertion into the plasma membrane, proteins destined to reside in the polar septum know that they've arrived at their correct address when they can reach out and touch the forespore.

References

1. Blobel, G. (1980). Intracellular protein topogenesis. Proc. Natl. Acad. Sci. USA *77*, 1496–1500.
2. Blaylock, B., Jiang, X., Rubio, A., Moran, C.P., Jr., and Pogliano, K. (2004). Zipper-like interaction between proteins in adjacent daughter cells mediates protein localization. Genes Dev. *18*, 2916–2928.
3. Doan, T., Marquis, K.A., and Rudner, D.Z. (2005). Subcellular localization of a sporulation membrane protein is achieved through a network of interactions along and across the septum. Mol. Microbiol., in press.
4. Stragier, P., and Losick, R. (1996). Molecular genetics of sporulation in Bacillus subtilis. Annu. Rev. Genet. *30*, 297–341.
5. Losick, R., Youngman, P., and Piggot, P.J. (1986). Genetics of endospore formation in *Bacillus subtilis*. Annu. Rev. Genet. *20*, 625–669.
6. Rudner, D.Z., and Losick, R. (2002). A sporulation membrane protein tethers the pro-sigmaK processing enzyme to its inhibitor and dictates its subcellular localization. Genes Dev. *16*, 1007–1018.
7. van Ooij, C., Eichenberger, P., and Losick, R. (2004). Dynamic patterns of subcellular protein localization during spore coat morphogenesis in Bacillus subtilis. J. Bacteriol. *186*, 4441–4448.
8. van Ooij, C., and Losick, R. (2003). Subcellular localization of a small sporulation protein in Bacillus subtilis. J. Bacteriol. *185*, 1391–1398.
9. Rudner, D.Z., Pan, Q., and Losick, R.M. (2002). Evidence that subcellular localization of a bacterial membrane protein is achieved by diffusion and capture. Proc. Natl. Acad. Sci. USA *99*, 8701–8706.
10. Londono-Vallejo, J.A., Frehel, C., and Stragier, P. (1997). SpoIIQ, a forespore-expressed gene required for engulfment in Bacillus subtilis. Mol. Microbiol. *24*, 29–39.
11. Rubio, A., and Pogliano, K. (2004). Septal localization of forespore membrane proteins during engulfment in Bacillus subtilis. EMBO J. *23*, 1636–1646.

Department of Molecular and Cellular Biology,The Biological Laboratories,16 Divinity Avenue, Harvard University, Cambridge, Massachusetts 02138, USA.
E-mail: losick@mcb.harvard.edu.

DOI: 10.1016/j.cub.2005.02.041

Virology: Gulliver among the Lilliputians

The discovery and genome sequencing of the mimivirus, a parasite of *Acanthamoeba*, blurs the boundary between viruses and cells: the 1.2 Mb genome of the mimivirus is predicted to contain 1262 genes and is much bigger than the genomes of many parasitic bacteria.

Eugene V. Koonin

The major discoveries of modern biology have come mostly through detailed molecular studies and comparative genomics. It is not common anymore, as it used to be in the 18th and 19th centuries, to discover marvelous creatures no one has ever seen before. Of course, in virology, which by definition deals with tiny intracellular parasites, the era of descriptive discoveries was delayed until the 20th century, and reports of new, sometimes unusual families of viruses continued into the new millennium [1]. Even so, the recent discovery [2] of the mimivirus, a parasite of the protozoan *Acanthamoeba polyphaga*, was entirely unexpected.

The mimivirus, the genome sequence of which has now been reported by Raoult *et al.* [3], is a true giant among viruses. Most strikingly, mimivirus crosses the boundary between viruses and cells that was considered more or less self-evident: viruses are assumed to be tiny and to have (much) smaller genomes than cellular life forms. At 1.2 Mb and with an estimated 1262 genes, the mimivirus genome is larger than the genomes of numerous parasitic bacteria and the single known parasitic archaeon, and only slightly smaller than the genomes of the simplest free-living prokaryotes (Figure 1).

The mimivirus genome has about 2.5 times as many genes as the smallest known prokaryotic genomes, those of the bacterium *Mycoplasma genitalium* and the archaeon *Nanoarchaeon equitans*. So it does not just nudge up to the virus—cell boundary, it leaps right across it. The physical dimensions of the virion are equally impressive: the icosahedral capsid of the mimivirus is at least 400 nm in diameter, about the same size as a small bacterial cell such as *Mycoplasma* [3].

These are the dramatic numbers, but what about the actual genetic content of the giant virus genome? The first thing to note is that, despite careful computational analysis, Raoult *et al.* [3] were able to assign homology-based functions to only 298 of the 1262 predicted genes (less than 25%). Most likely, extensive searches for subtle sequence and structural similarities will lead to additional functional assignments, but the current numbers are notably different from the typical results of analysing newly sequenced prokaryotic genomes. These days, at least for smaller bacterial and archaeal genomes, about 70% of the predicted genes have homologs with known functions [4].

Compared to prokaryotic genomes, therefore, the similar-sized genome of the mimivirus is almost like *terra incognita*. However, analysis of the evolutionary affinities and predicted functions of those genes that do have well-characterized homologs clearly shows that mimivirus did not originate from Mars, but has a lot in common with other viruses. These genes can be classified into two major categories: genes shared with all or some nucleocytoplasmic large DNA viruses (NCLDVs); and genes with prokaryotic and/or eukaryotic homologs not represented in other NCLDVs.

Earlier comparative analysis showed that the NCLDVs — which include poxviruses, iridoviruses, asfarviruses and phycodnaviruses — share a core set of conserved

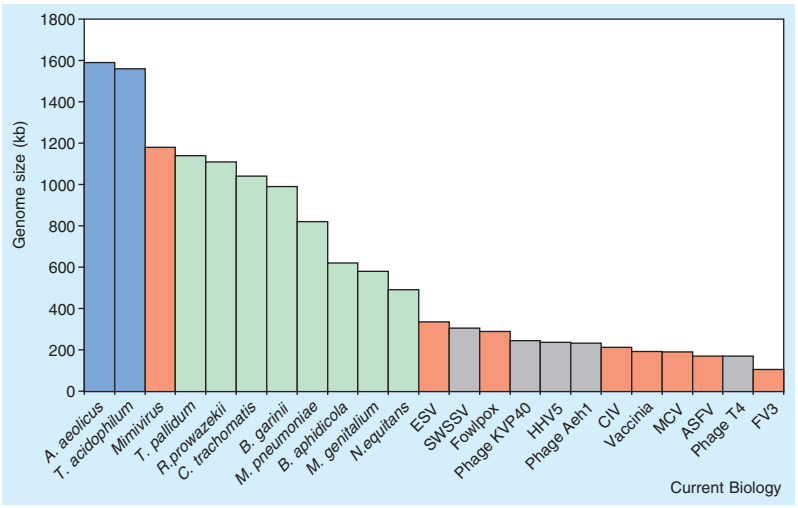

Figure. 1. Genome sizes of selected double-stranded DNA viruses, bacteria and archaea.

Dark blue, free-living prokaryotes; light green, parasitic prokaryotes; red, NCLDV; gray, other dsDNA viruses. Abbreviations: A. aeolicus, *Aquifex aeolicus*; T. acidophilum, *Thermoplasma acidophilum*; T. pallidum, *Treponema pallidum*; R. prowazekii, *Rickettsia prowazekii*; C. trachomatis, *Chlamydia trachomatis;*, B. garinii, *Borrelia garinii*; M. pneumoniae, *Mycoplasma pneumoniae*; B. aphidicola, *Buchnera aphidicola*; M. genitalium, *Mycoplasma genitalium*; N. equitans, *Nanoarchaeum equitans*; ESV, Ectocarpus siliculosus virus (phycodnavirus); SWSSV, Shrimp white spot syndrome virus; HHV5, human herpesvirsus 5; CIV, Chilo iridescent virus (iridovirus); MCV, Molluscum Contagiosum virus (poxvirus); ASFV, African swine fever virus (asfarvirus); FV3, Frog virus 3 (iridovirus). The data are taken from the genome division of the Entrez database at the National Center for Biotechnology Information (NIH, Bethesda): http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome.

genes, implying that these diverse viruses originated from a common ancestor [5]. Evolutionary reconstructions based on the parsimony principle assigned 31 genes to this hypothetical ancestral virus. The products of these genes are responsible, largely, for viral genome replication and transcription, but also for virion biogenesis. Of these ancestral genes, the mimivirus lacks only six, leaving no doubt that, its colossal size notwithstanding, it shares a common origin with other NCLDVs (see Supplementary Table 1S at ftp://ftp.ncbi.nih.gov/pub/koonin/mimivirus). The mimivirus proteins do not show unequivocal affinity to homologs from any of the other lineages of NCLDV, so it definitely represents a new branch of this viral class.

The mimivirus genes with non-viral homologs are a diverse and remarkable group; many encode proteins not previously seen in viruses. Among these are: several translation system components, in particular, four aminoacyl-tRNA synthetases, four translation factors, a tRNA-modification enzyme and several tRNAs; enzymes involved in biosynthesis of amino acids, in particular, glutamine and asparagine; several DNA repair enzymes; and several proteins with molecular chaperone functions (Table 1S).

In itself, the presence of genes with these functions in a viral genome is not unprecedented. Thus, repair genes are present in all NCLDV genomes, and PBCV-1 encodes a translation elongation factor [6]; furthermore, some bacteriophages encode tRNAs [7], while molecular chaperones are encoded by the largest RNA viruses of plants, the closteroviruses [8]. But the diversity of mimivirus's repertoire of central cellular functions, particularly, translation, far exceeds that observed in other viruses and comes as a real surprise. This may reflect an infection strategy that involves deep reprogramming of cellular functions, particularly translation, presumably enhancing expression of viral genes by mechanisms that remain to be investigated.

What is the origin of the 'cellular' genes of the mimivirus? A quick analysis shows that this group is heterogeneous. The translation-associated proteins resemble the eukaryotic counterparts, whereas most of the repair genes are more closely related to bacterial orthologs; the other enzymes also show either eukaryotic or bacterial affinities (Table 1S). In some cases, such as that of DNA ligase, the ancestral NCLDV enzyme (ATP-dependent ligase) has apparently been displaced by a bacterial enzyme (NAD-dependent ligase); notably, the same displacement occurred independently in entomopoxviruses [9].

An excellent example of the diverse evolutionary histories of mimivirus genes is presented by the three distinct topoisomerases: topoisomerase IB, which is also found in all poxviruses, though the mimivirus protein is much more similar to bacterial than to poxvirus orthologs; topoisomerase IIA, which is highly similar to orthologs from PBCV1 and eukaryotes; and, topoisomerase IA, which is highly similar to bacterial homologs and has so far not been seen in viruses.

Obviously, little can be said at this point about the functions and origins of the more than 900 genes of the mimivirus that do not have readily detectable homologs. But hints come from searching for conserved domains and comparison among the mimivirus proteins themselves. Over 30 mimivirus proteins contain the ankyrin repeats, and more than 20 contain POZ (BTB) domains [3]. Both types of domain are known to be involved in protein–protein interactions and the formation of macromolecular complexes [10], suggesting that the mimivirus encodes a complex apparatus for virion morphogenesis and intracellular transport. Furthermore, comparison of mimivirus protein sequences reveals many families of uncharacterized paralogs, some with more than 20 members (my unpublished observations). So gene duplication clearly played a major role in mimivirus evolution.

The enormous complexity of the mimivirus genome — for a virus — makes its origin and evolutionary history particularly intriguing. Raoult *et al.* [3] propose two non-trivial hypotheses: that mimivirus evolved by genome degradation from an even more complex entity which encoded, among other systems, a nearly complete translation machinery; and that the mimivirus lineage diverged from other cellular life forms very early in evolution, perhaps before the divergence of bacteria, archaea, and eukaryotes (this is inferred from the topology of a multigene tree in which the mimivirus branch does not show clear affinity to any of the cellular lineages).

I believe that neither of these suggestions is supported by the results of mimivirus genome analysis. On the contrary, a number of observations suggest that the vast gene repertoire of the mimivirus evolved by piecemeal accretion of genes on top of the ancestral NCLDV core. These are: the way that the mimivirus shares the core set of genes with the other NCLDVs; the different (eukaryotic and bacterial) evolutionary affinities of the non-viral genes in the mimivirus genome; and the presence of many paralogous gene families.

The ancestors of mimivirus apparently derived genes from the eukaryotic hosts, endosymbiotic bacteria and, possibly, other viruses. The mimivirus genome further grew through extensive gene duplication. Mimivirus genome inflation might have been driven by evolutionary processes similar to those that shaped the increased genomic complexity of cellular life forms [11,12]. Virus–host interactions in Protozoa might have resulted in the virus population having a low effective size, weakening purifying selection and increasing the likelihood of acquisition of foreign genes and fixation of gene duplications.

The topology of the multigene tree is not a strong argument for the mimivirus lineage having an ancient origin, given the propensity of viral genes for fast evolution. On the contrary, the unique combination of about 30 conserved NCLDV genes is not seen outside eukaryotic viruses. It is therefore unlikely that the NCLDVs are older than eukaryotes, and the mimivirus lineage should have diverged from the common ancestor with other viruses even later. It is a distinct possibility, however, that the NCLDV ancestor existed at a very early stage of eukaryotic evolution, before the radiation of the major lineages of eukaryotes.

Even if the most radical ideas on mimivirus evolution are poorly compatible with the data, the discovery [2] and genome sequencing [3] of this virus open a new chapter in virus genomics. The conservation of the NCLDV gene core and the accretion of a huge number of additional genes attest to the incredible plasticity of viral genomes. For all its gigantic dimensions and unusual gene repertoire, the mimivirus is a *bona fide* virus, which depends on the host cell for translation of its proteins and for energy, and has a typical capsid, proving that the difference between viruses and cells is indeed one in kind not just in degree. It remains to be seen whether or not the mimivirus is a true record-holder among viruses (unlikely) and how common are such giant viruses in unicellular eukaryotes (there could be scores of them).

References

1. Fields, B.N., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E., and Knipe, D.M. (2001). Fields Virology (New York: Lippincott Williams & Wilkins).
2. La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M., and Raoult, D. (2003). A giant virus in amoebae. Science *299*, 2033.
3. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.M. (2004). The 1.2-megabase genome sequence of Mimivirus. Science *306*, 1344–1350.
4. Galperin, M.Y., and Koonin, E.V. (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. *32*, 5452–5463.
5. Iyer, L.M., Aravind, L., and Koonin, E.V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. *75*, 11720–11734.
6. Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L., and Van Etten, J.L. (1997). Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. Virology *237*, 360–377.
7. Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003). Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. *67*, 86–156.
8. Peremyslov, V.V., Hagiwara, Y., and Dolja, V.V. (1999). HSP70 homolog functions in cell-to-cell movement of a plant virus. Proc. Natl. Acad. Sci. USA *96*, 14771–14776.
9. Afonso, C.L., Tulman, E.R., Lu, Z., Oma, E., Kutish, G.F., and Rock, D.L. (1999). The genome of Melanoplus sanguinipes entomopoxvirus. J. Virol. *73*, 533–552.
10. Pawson, T., Raina, M., and Nash, P. (2002). Interaction domains: from simple binding events to complex cellular behavior. FEBS Lett. *513*, 2–10.
11. Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science *302*, 1401–1404.
12. Koonin, E.V. (2004). A non-adaptationist perspective on evolution of genomic complexity or the continued dethroning of man. Cell Cycle *3*, 280–285.

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA.
E-mail: koonin@ncbi.nlm.nih.gov

DOI: 10.1016/j.cub.2005.02.042

Polytopic Proteins: Preventing Aggregation in the Membrane

It has been proposed that the aggregation of nascent transmembrane segments of polytopic proteins is prevented by chaperones present in the endoplasmic reticulum membrane; now the first experimental support for this proposal has been reported.

J. Michael Lord[1] and Stephen High[2]

It is generally accepted that 'polytopic' membrane proteins — the polypeptide chains of which cross the membrane multiple times — are integrated at the classical Sec61-based endoplasmic reticulum (ER) translocon [1], but the mechanism by which the translocon deals with multiple transmembrane domains is less clear. Two models have been proposed to describe this process. One is the '*en bloc*' model, in which all the